Constrained Optimization II: Penalty Method

Asst. Prof. Dr.-Ing. Sudchai Boonto

November 12, 2024

Department of Control System and Instrumentation Engineering
King Mongkut's Unniversity of Technology Thonburi
Thailand

Objective

At the end of this chapter you should be able to:

- Penalty methods
    - Quadratic Penalty Method
    - Augmented Lagrange Method
    - Interior Penalty Method

Exterior Penalty Methods

There are two popular penalty method:

- **Quadratic penalty:** The quadratic penalties are continuously differentiable and straightforward to implement, but they suffer from numerical ill-conditioning.
- **Nonsmooth exact penalty:** This method uses a single unconstrained problem to take palce the original constrained problem. Using these penalty functions, we can often find a solution by performing a single unconstrained minimization, buth the nonsmoothness may create complications. A popular function of this type is the ℓ_1 penalty function.
- **Augmented Lagrangian:** The augmented Lagrangian method is more sophisticated; it is based on the quadratic penalty but adds terms that improve the numerical properties.
- **Interior penalty:** Similar to quadratic penalty method, but working in opposite way.

The Quadratic Penalty Method

We want to replace a constrained optimization problem by a single function consisting of

- the original objective function, and
- additional term for each constraint, which is positive when the current point $\mathbf{x}$ violates the constraint and zero otherwise.

The method defines the **sequence** of such penalty functions, in which the penalty terms for the constraint violations are multiplied by a positive coefficient.

- By making this coefficient larger, we penalize constraint violations more severely
- thereby forcing the minimizer of the penalty function closer to the feasible region for the constrained problem.
- The simplest penalty function of this type is the **quadratic penalty function**, in which the penalty terms are the squares of the constraint violations.

The Quadratic Penalty Method

Consider the equality constrained problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p \end{aligned}$$

The quadratic penalty function $Q(\mathbf{x}; \mu)$ for this formulation is

$$Q(\mathbf{x}; \boldsymbol{\mu}) = f(\mathbf{x}) + \frac{\mu}{2} \sum_{i=1}^{p} h_i^2(\mathbf{x}),$$

where $\mu > 0$ is the **penalty parameter.**

- By driving μ to ∞, we penalize the constraint violations with increasing severity.
- We use a sequence of values $\{\mu_k\}$ with $\mu_k \uparrow$ as $k \to \infty$, and to seek the approximate minimizer $\mathbf{x}_k$ of $Q(\mathbf{x}; \mu_k)$ for each k.

The Quadratic Penalty Method: Example

Consider a problem

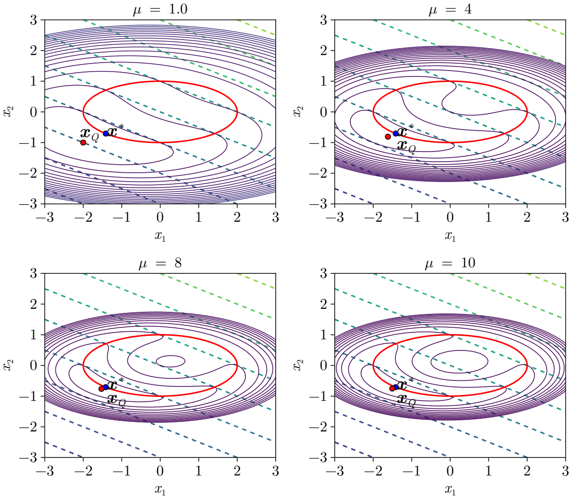
$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & x_1 + 2x_2 \\ \text{subject to} \quad & \frac{1}{4}x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

for which the solution is $\begin{bmatrix}-1.414 & -0.707\end{bmatrix}^T$ and the quadratic penalty function is

$$Q(\mathbf{x};\mu) = x_1 + 2x_2 + \frac{\mu}{2}\left(\frac{1}{4}x_1^2 + x_2^2 - 1\right)^2$$

When $\mu = 1$, we observe a minimizer of Q near the point $\begin{bmatrix}-2 & -1\end{bmatrix}^T$. The penalty is active for all points that are infeasible, but the minimum of the penalized function does not coincide with the constrained minimum of the **original constrained problem**. The penalty parameter needs to be increased for the minimum of th epenalized function to approach the correct solution, but this results in a poorly conditioned function.

The Quadratic Penalty Method: Example



Here, the red line is the constraint, the dashed line is the objective function, and the primary contour is the contour of the Q function.

The Quadratic Penalty Method: Algorithm

Require:

$\mathbf{x}_0$: Starting point

$\mu_0 > 0$: Initial penalty parameter

$\rho > 1$: Penalty increase factor ($\rho \sim 1.2$ is conservative, $\rho \sim 10$ is aggressive)

$k = 1$

while $\|\nabla_{\mathbf{x}} Q(\mathbf{x}; \mu_k)\| \leq \varepsilon$ **do**

  $\mathbf{x}_k^* \leftarrow \text{minimize}_{\mathbf{x}_k} Q(\mathbf{x}; \mu_k)$

  $\mu_{k+1} = \rho \mu_k$

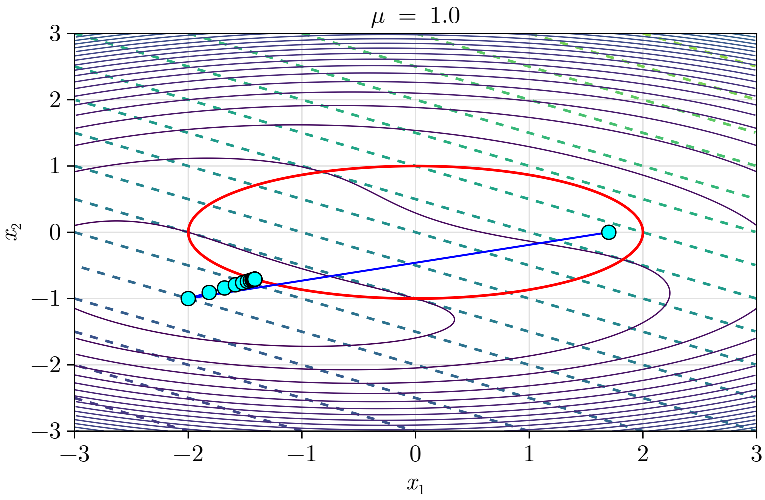
  $\mathbf{x}_{k+1} = \mathbf{x}_k^*$

  $k = k + 1$

end while

For minimization, we can use any method explain in unconstrained problem.

The Quadratic Penalty Method: Algorithm



Using a previous algorithm with the starting $\mu = 1.0$.

The Quadratic Penalty Method: Example

For a given value of the penalty parameter μ, the penalty function may be unbounded below een if the original consrained problem has a unique solution. Consider for example

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -5x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1 = 1 \end{aligned}$$

- The solution is at $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$
- The penalty function is unbounded for any $\mu < 0$.
- For such values of μ, the iterates generated by an unconstrained minimization method would usually diverge.
- This deficiency is common to all the penalty function.

The Quadratic Penalty Method: General formula

For the general constrained optimization problem

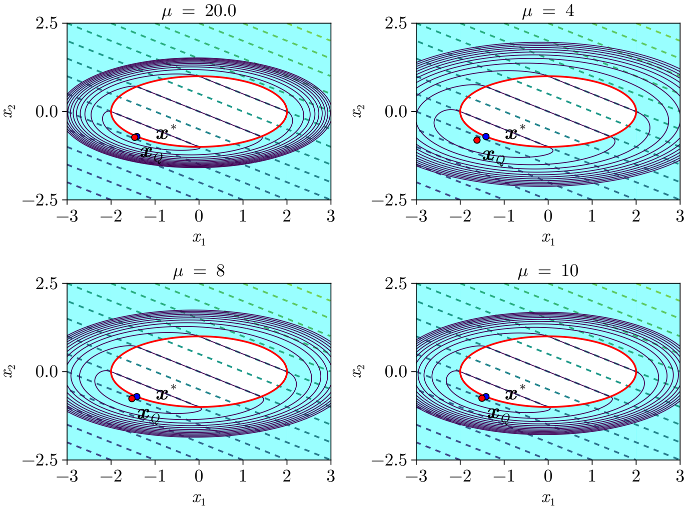
$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \\ & g_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, q \end{aligned}$$

We can define the quadratic penalty function as

$$Q(\mathbf{x}; \mu) = f(\mathbf{x}) + \frac{\mu_h}{2} \sum_{i=1}^{p} h_i^2(\mathbf{x}) + \frac{\mu_g}{2} \sum_{j=1}^{q} \max(0, g_i(\mathbf{x}))^2$$

The last term means that we add the penalty wehn the inequality constraint is violated (i.e., when $g_i(\mathbf{x}) > 0$).

The Quadratic Penalty Method: Example



Here, the red line is the constraint, the dashed line is the objective function, and the primary contour is the contour of the Q function. The white area is the feasible region.

The issues of the penalty method

- If the starting value for μ is too low. The penalty might not be enough to overcome a function that is unbounded from below, and the penalized function has no minimum.
- In practical, we can not make $\mu \to \infty$. Hence, the solution to the problem is always slightly infeasible. By comparing,

$$\nabla_{\mathbf{x}}\mathcal{L} = \nabla f + \mathbf{J}_{\mathbf{h}}^T \boldsymbol{\lambda} = 0 \qquad \text{the optimality condition of the constrained problem}$$

$$\nabla_{\mathbf{x}} Q = \nabla f + \mu \mathbf{J}_{\mathbf{h}}^T \mathbf{h} = 0 \qquad \text{the optimality condition of the penalized function.}$$

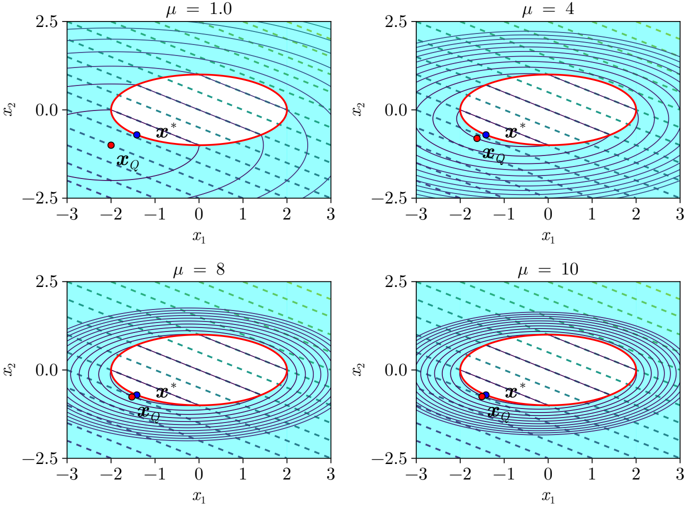
$$h_i \approx \frac{\lambda_j^*}{\mu}$$

  Because $h_j = 0$ at the optimum, μ must be large to satisfy the constraints.
- The extra curvature is added in a direction perpendicular to the constraints, making the Hessian of the penalized function increasingly ill-conditioned as μ increases. Thus, the need to increase μ to improve accuracy directly leads to a function space that is increasingly challenging to solve.

The ℓ_1-norm Penalty Method: Example

The ℓ_1 **penalty function** is defined by

$$Q(\mathbf{x};\mu) = f(\mathbf{x}) + \mu_h \sum_{i=1}^{p} |h_i(\mathbf{x})| + \mu_g \sum_{j=1}^{q} |\max(0, g_i(\mathbf{x}))|$$



Augmented Lagrangian method

- The next method is The **method of multipliers** or the **augmented Lagrangian** method. This algorithm is related to the quadratic penalty algorithm, however it reduces the possibility of ill conditioning by introducing explicit Lagrange multiplier estimates into the function to be minimized, which is known as the augmented Lagrangian function.
- In contrast to the penalty functions, the augmented Lagrangian function largely preserves smoothness, and implementations can be constructed from standard software for unconstrained or bound-constrained optimization.

The augmented Lagrangian function for equality constraints is

$$\mathcal{L}_A(\mathbf{x};\boldsymbol{\lambda},\mu) = f(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i h_i(\mathbf{x}) + \frac{\mu}{2}\sum_{j=1}^{q} h_j(\mathbf{x})^2$$

Augmented Lagrangian method

To estimate the Lagrange multipliers, we can compare the optimality conditions for the augmeted Lagrangian and the actual Lagrangian,

$$\nabla_{\mathbf{x}} \mathcal{L}_A(\mathbf{x}; \boldsymbol{\lambda}, \mu) = \nabla f(\mathbf{x}) + \sum_{i=1}^{p} (\lambda_i + \mu h_i(\mathbf{x})) \nabla h_j = 0$$

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_j(\mathbf{x}^*) = 0$$

We have

$$\lambda_i^* = \lambda_i + \mu h_i \quad \Rightarrow \quad \boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \mu_k h(\mathbf{x}_k)$$

This method can improve the plain quadratic penalty because updating the Lagrange multiplier estimates at each iteration allows for more accurate solutions without increasing μ as much. Since

$$h_i \approx \frac{1}{\mu}(\lambda_i^* - \lambda_i) \quad \text{compare to the quadratic penalty method} \quad h_i \approx \frac{\lambda_i^*}{\mu}$$

Augmented Lagrangian method

- The quadratic penalty relies solely on increasing μ in the denominator to drive the constraints to zero.
- The augmented Lagrangian also controls the numerator through the Lagrange multiplier estimate. If the estimate is reasonable close to the true Lagrange multiplier, then the numerator becomes small for modest values of μ.
- Thus, the augmented Lagrangian can provide a good solution for $\mathbf{x}^*$ while avoiding the ill-conditioning issues of the quadratic penalty.

The the general augmented Lagrangian including the inequality constraints we have

$$\mathcal{L}_A(\mathbf{x}; \boldsymbol{\lambda}, \mu) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \bar{g}(\mathbf{x}) + \frac{\mu}{2} \|\bar{g}(\mathbf{x})\|_2^2,$$

where

$$\bar{g}_i(\mathbf{x}) = \begin{cases} h_i(\mathbf{x}) & \text{for equality constraints} \\ g_i(\mathbf{x}) & \text{if } g_i \geq -\lambda_i/\mu \\ -\lambda_i/\mu & \text{otherwise} \end{cases}$$

The augmented Lagrange Method: Algorithm

Require:

$\mathbf{x}_0$: Starting point

$\boldsymbol{\lambda}_0 = 0$ Initial Lagrange multiplier

$\mu_0 > 0$: Initial penalty parameter

$\rho > 1$: Penalty increase factor

$k = 0$

while $\|\nabla_{\mathbf{x}} \mathcal{L}_A(\mathbf{x}; \boldsymbol{\lambda}, \mu_k)\| \geq \varepsilon$ **do**

$\quad \mathbf{x}_k^* \leftarrow \text{minimize}_{\mathbf{x}_k} \ \mathcal{L}_A(\mathbf{x}; \boldsymbol{\lambda}_k, \mu_k)$

$\quad \boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \mu_k \mathbf{h}(\mathbf{x}_k)$

$\quad \mu_{k+1} = \rho \mu_k$

$\quad \mathbf{x}_{k+1} = \mathbf{x}_k^*$

$\quad k = k + 1$

end while

For minimization, we can use any method explain in unconstrained problem.

Augmented Lagrangian method: Example

Consider

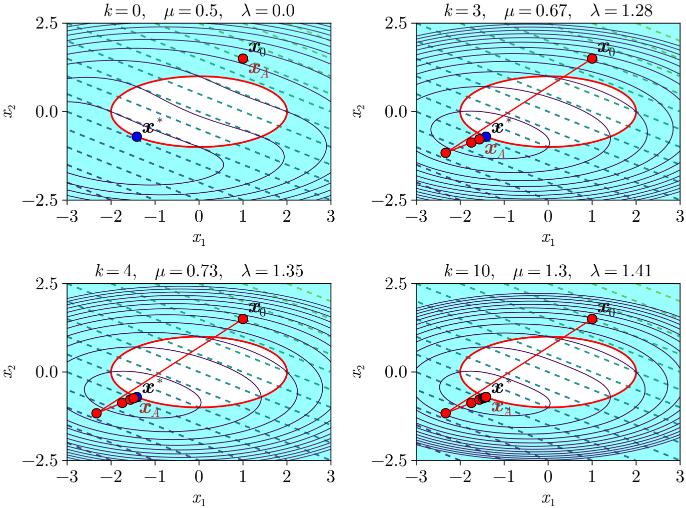
$$\begin{aligned}
&\text{minimize} \quad f(\mathbf{x}) = x_1 + 2x_2 \\
&\text{subject to} \quad g(\mathbf{x}) = \frac{1}{4}x_1^2 + x_2^2 - 1 \leq 0
\end{aligned}$$

The augmented Lagrangian is

$$\mathcal{L}_A(\mathbf{x}; \mu) = x_1 + 2x_2 + \lambda \left(\frac{1}{4}x_1^2 + x_2^2 - 1 \right) + \frac{\mu}{2} \left(\frac{1}{4}x_1^2 + x_2^2 - 1 \right)^2$$

Applying the augmented Lagrangian algorithm, starting with $\mu = 0.5$ and using $\rho = 1.1$.

Augmented Lagrangian method: Example



Compared with the quadratic penalty, the penalized function is much better conditioned. Th eminimum of the penalized function eventually becomes the minimum of the constrained problem without a large penalty parameter.

Interior penalty method

The interior penalty method is they seek to maintain feasibility.

- Instead of adding a penalty only when constraints are violated, they add a penalty as the constraint is approached from the feasible region.
- This type of penalty is particularly desirable if the objective function is ill-defined outside the feasible region.
- The methods are call **interior** because the iteration points remain on the interior of the feasible region.
- They are also referred to as **barrier methods** because the penalty function acts as a barrier preventing iterates from leaving the feasible region.
- Theinterior penalty function to enforce $g(\mathbf{x}) \leq 0$ is the **inverse barrier**

$$\pi(\mathbf{x}) = \sum_{j=1}^{q} -\frac{1}{g_j(\mathbf{x})},$$

where $\pi(\mathbf{x}) \to \infty$ as $g_j(\mathbf{x}) \to 0^-$

Interior penalty method

- A more popular interior penalty function is the **logarithmic barrier**

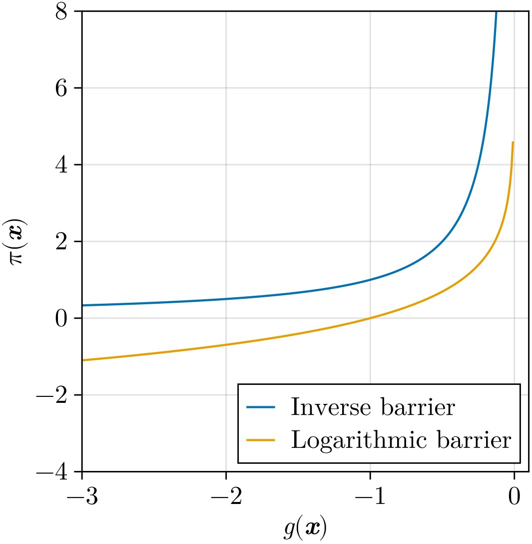
$$\pi(\mathbf{x}) = \begin{cases} \sum_{j=1}^{q} -\ln\left(-g_j(\mathbf{x})\right) & g_i(\mathbf{x}) < 0 \\ 0 & \text{otherwise} \end{cases}$$

  which also approaches infinity as the constraint tends to zero from the feasible side. The penalty function is then

$$Q_b(\mathbf{x}; \mu) = f(\mathbf{x}) + \mu\pi(\mathbf{x})$$

- Neither of these penalty functions applies when $g_j > 0$ because they are designed to be evaluated only within the feasible space. Algorithms based on these penalties must be prevented from evaluating infeasible points.

Interior penalty method



The Interior Penalty Method: Algorithm

Require:
$\mathbf{x}_0$: Starting point
$\mu_0 > 0$: Initial penalty parameter
$\rho < 1$: Penalty increase factor
$k = 1$
while $\|\nabla_{\mathbf{x}} Q_b(\mathbf{x}; \mu_k)\| \geq \varepsilon$ **do**
  $\mathbf{x}_k^* \leftarrow \text{minimize}_{\mathbf{x}_k}\, Q_b(\mathbf{x}; \mu_k)$
  $\mu_{k+1} = \rho \mu_k$ Decrease penalty parameter
  $\mathbf{x}_{k+1} = \mathbf{x}_k^*$
  $k = k + 1$
end while

For minimization, we can use any method explain in unconstrained problem.

The Interior Penalty Method: Example

Consider a problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & x_1 + 2x_2 \\ \text{subject to} \quad & \frac{1}{4}x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

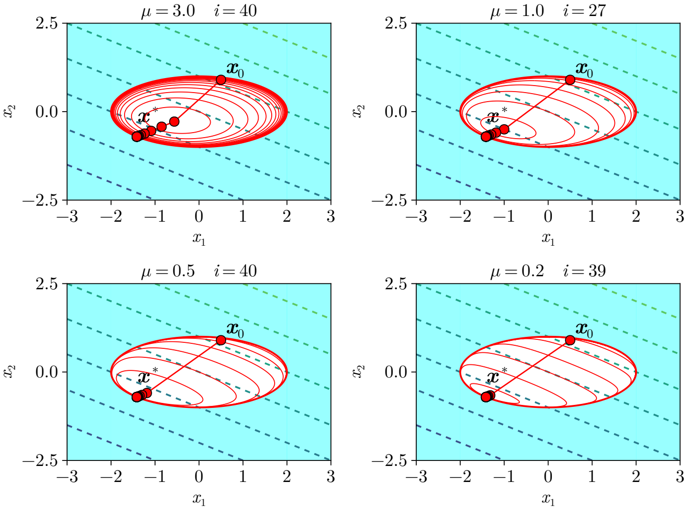
for which the solution is $\begin{bmatrix} -1.414 & -0.707 \end{bmatrix}^T$ and the logarithmic penalty function is

$$Q_b(\mathbf{x}; \mu) = x_1 + 2x_2 - \mu \ln\left(-\frac{1}{4}x_1^2 - x_2^2 + 1\right)$$

Note:

- the Hessian for interior penalty methods becomes increasingly ill-conditioned as the penalty parameter tends to zero.
- There is a modern method that can solve the ill-conditioned problem such at interior-point method.

The Interior Penalty Method: Example



Reference

1. Joaquim R. R. A. Martins, Andrew Ning, ”*Engineering Design Optimization*,” Cambridge University Press, 2021.
2. Mykel J. kochenderfer, and Tim A. Wheeler, ”*Algorithms for Optimization*,” The MIT Press, 2019.
3. Jorge Nocedal, and Stephen J. Wright, ”*Numerical Optimization*,” 2nd, Springer, 2016
4. Stephen Boyd, and Lieven Vandenberghe , ”*Convex Optimization*,” Cambridge University Press, 2009.